

本周周报（2013.11.04-2013.11.10）

郭方舟

本周工作

1. 空气污染数据可视化

本周主要实现了 ck 距离的计算，以及基于 ck 距离的一些基本交互。

1) Ck 距离的计算

CK 距离计算通过调用 matlab 代码实现，CK 距离实际上是使用两张图片生成 mpeg 文件的大小来计算的，其公式是：

$$d_{mpeg}(x, y) = \frac{C(x|y) + C(y|x)}{C(x|x) + C(y|y)} - 1$$

其中，x 和 y 就是需要比较的两张图片，C(x|y)是指 x 作为第一帧，y 作为第二帧所生成的 mpeg 文件的大小。

在调用 matlab 代码的过程中，发现 matlab 的数据访问方式和 java 的数据访问方式不同，因此在两者的数据通信过程中遇到了一些麻烦，主要就是三维数组的组织方式不同导致传给 matlab 的图片数据无法正确读取。

2) 选择时间点和窗长进行 ck 距离计算

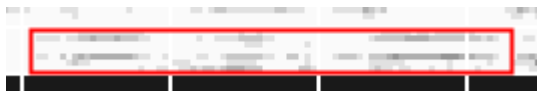
这个交互目前是通过在熵图上进行右键点选实现的。在熵图上通过左键点击并发现像素图中的一些模式之后，右键点击选定的点，将计算选定属性在所有时间点上，以选定窗长为长度的片段与选定的窗的 ck 距离。

3) 计算结果的显示

得到的所有 ck 距离会在熵图下方显示出来，每个值以 10*10 的方块表示，灰度越高表示 ck 距离越小，也就是与选定的片段越相似：

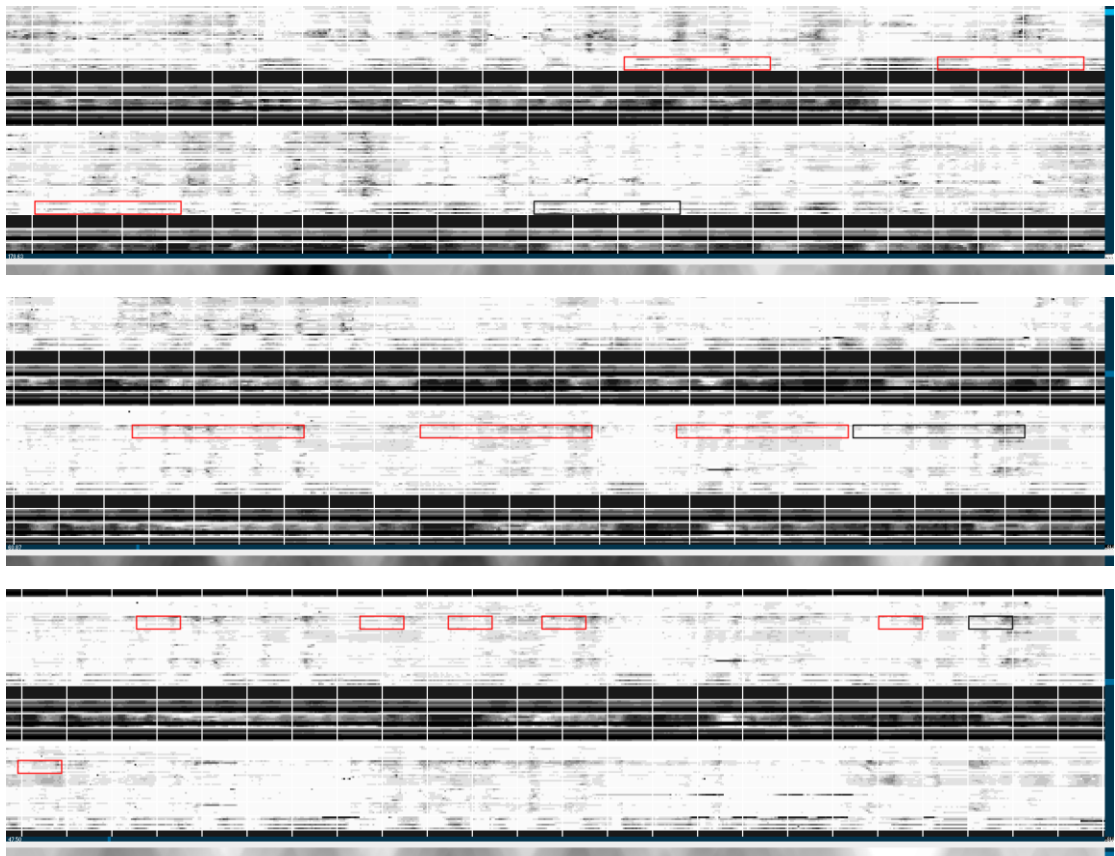


在颜色带上点击，会在像素图上以红色矩形标出点击的像素块所代表的位置：



4) 在像素图上标出最相似的十个时间点

由于计算出的 CK 距离有八千多个，为了方便起见，将 ck 距离最小的 10 个时间点。效果如图所示：



在标出 top 10 的时候，需要过滤掉两个片段有重合的情况，否则标出的 top 10 就会集中在某几个片段上。

从上面给出的结果可以看到，使用 ck 距离能够找出与选定的片段相似的片段。

下周工作

空气污染数据的可视化

学习 transform entropy 算法，这个算法有一个 matlab 的实现，但是这个实现是针对脉冲数据的，因此只能计算 0,1 串，对我们的数据不适用。如果想要使用这个算法计算两个属性或是两个事件之间的关系的话，接下来就要实现这个算法。